A Cross-Country Analysis of the Relationship Between National Income and Life Expectancy

Patrick Taylor

ECON 3161

Georgia Institute of Technology

April 30, 2021

Abstract

International objectives for increasing life expectancy and overall health have generally been focused on increasing the quality of healthcare accessible to individuals. These initiatives for increasing the quality of healthcare almost always pertain to reducing healthcare expenses, making it easier and more affordable for individuals to receive the healthcare that they need. Evidence suggests that social and economic factors such as income and level of education may have a larger effect on increasing general well-being and life expectancy than previously thought.

Introduction

Life expectancy has risen dramatically over the past few centuries, with the global average life expectancy having more than doubled since 1900. The inequality of life expectancy is still apparent, however, when examining data across a large number of countries. The goal of this paper is to examine and identify the factors that can lead to an increase in life expectancy, as well as to compare these factors across a number of different countries.

As life expectancy began to rise throughout industrialized countries in the early 19th century, it stayed low across the rest of the world. Instinctively, it could be concluded that richer countries had access to better healthcare and the income to support healthier, more relaxed lifestyles, which in turn led to a higher life expectancy amongst its citizens. In comparison, citizens of poorer countries lacked the resources and infrastructure that act as leading factors in increasing life expectancy. Over the past several decades, this inequality has decreased significantly, with no country having a lower life expectancy than the country with the highest life expectancy in 1800. There is, however, still a large disparity between the life expectancy of citizens in wealthy, developed countries and those from poorer countries. This paper plans to examine the relationship between national income, as measured through GDP per capita, and the life expectancy of countries across the globe. We expect to see a strong relationship between national income and life expectancy, as wealthier countries likely have better healthcare infrastructure and less stressful lifestyles than poorer countries, both of which have been identified as factors that can lead to higher life expectancy.

Additionally, we plan to examine the relationship between education spending, healthcare spending, population growth and life expectancy across a number of different countries. Similar to income, we would expect there to be a positive correlation between a country's education spending and the life expectancy of its citizens. Previous studies have shown that highly educated adults in the U.S. have annual mortality rates lower than less-educated adults, so we would expect to see countries that spend more on education have a higher life expectancy of its citizens than countries that spend less on education.

Our hypothesis is that there is a strong correlation between national income and life expectancy. Citizens of countries with a higher level of national income likely have a higher standard of living, healthier lifestyles, and access to better healthcare infrastructure, all of which contribute to a longer life expectancy.

Literature Review

Numerous studies have established a relationship between income and life expectancy, however, this relationship is still not completely understood. Chetty, Stepner, Cutler (2016) examined the association between income and life expectancy in the United States from the year 2001 through 2014. The study found a correlation between income and life expectancy, with higher income being associated with a longer life expectancy. With a sample of 1,408,287,218 person-year observations, the study yielded four main conclusions about the relationship between income and life expectancy. The two conclusions that are most significant for the purposes of this research paper include higher income being associated with longer life expectancy and that inequality in life expectancy has increased over time. The study found that the gap in life expectancy between individuals in the top 1% of income and the bottom 1% of income was 14.6 years for men and 10.1 years for women. Additionally, the study found that between 2001 and 2014 life expectancy for men and women in the top 5% of the income distribution increased by 2.34 years and 2.91 years respectively, while the life expectancy for men and women in the bottom 5% of the income distribution increased by only .32 years and .04 years respectively. The study concluded that higher income was associated with longer life expectancy, however, the differences in life expectancy were also related to differences in healthcare infrastructure, health characteristics, and local area characteristics.

Chalhoub and Twomey (2018) outline a study that found that life expectancy increases as income increases, and it also found that the difference in life expectancy between the top and bottom quartiles of income varies by geographical area. Factors such as smoking, obesity, and exercise can explain some of the difference in life expectancy between the top and bottom quartiles of income. What's more, the study suggests that lower income individuals living in areas with a higher standard of living (higher home values, more college graduates, etc.) also practice healthier behaviors and will typically have a higher life expectancy than individuals with equally low income that live in low income areas. More generally, the study concludes that cities or areas with higher standards of living and healthier behaviors will also have a higher life expectancy for lower-income individuals. Despite this conclusion, the study further examines the multifaceted relationship between income and life expectancy by analyzing the contribution of factors such as smoking to the difference in life expectancy between individuals of different income levels.

Much like income level, numerous studies have found a correlation between education and life expectancy. Hernandez and Hummer (2013) found that remaining life expectancy at age 25 is nearly 10

years longer for adults that have a college education in comparison to those who do not have a high school diploma. The study was performed by analyzing mortality rates of adults over the age of 25 after splitting them into four groups by education level: less than a high school degree, a high school degree, some college but not a bachelor's degree, a bachelor's degree or higher. The main conclusion that the study drew was that individuals with a higher education level had significantly lower mortality rates than those with less education. One example of this is the difference in mortality rate between individuals with less than a high school degree and those with a bachelor's degree or more. Men with less than a high school degree have a mortality rate over 4 times higher than men with a bachelor's degree or higher and women with less than a high school degree have a mortality rate nearly 4 times higher than women with a bachelor's degree or higher. After analyzing the mortality rates of several different demographics and levels of education the study came to the conclusion that differences in education and mortality have widened significantly over the past 20 years. The benefits of a higher level of education include better occupations and higher levels of income, higher cognitive function, and healthier behavior. All of these factors can contribute to a longer life expectancy as a result of pursuing a higher level of education.

Rather than analyzing the relationship between individuals across a distribution of income levels and their life expectancy, this paper seeks to examine the relationship between national income and national life expectancy across a number of different countries. Given the relationship found by previous studies between individual income, education level, and life expectancy, our hypothesis is that countries with higher levels of national income (as measured by GDP per capita) and spending on education will have higher life expectancies. Citizens of countries that have a higher level of national income likely lead healthier lifestyles and have access to better healthcare infrastructure, both of which are contributing factors in life expectancy. Additionally, citizens of countries that spend more on education are unsurprisingly likely to be more highly educated. As examined in one study mentioned above, a higher education level can lead to better occupations, higher income, and a healthier lifestyle, all of which are correlated to a longer life expectancy.

Data

The main focus of this study is to analyze the relationship between GDP per capita and life expectancy at birth for a large group of countries. The dependent variable in our study is life expectancy at birth and our independent variables are GDP per capita, government expenditures on education, healthcare spending, and population growth. We started with data from nearly 250 countries and after dropping missing entries we were left with the data of 146 nations and regions from 2017. All of the data being used was collected from the World Bank.

We chose to examine the effect of national income and government education spending on national life expectancy based on the relationship found between them on an individual level in the studies mentioned above. When examining the data on a national level, we expect to see a positive correlation between national income and life expectancy, and perhaps an even stronger correlation between education spending and life expectancy given the numerous benefits that come from a higher level of education as mentioned previously. Additionally, we will examine the relationship between healthcare spending, population growth, and life expectancy, since instinctively one would assume these two variables have some effect on the life expectancy of a country's citizens.

We will use two models to test our hypothesis, one single variable regression model to examine the relationship between national income and life expectancy and a multiple regression model to account for national income, education spending, healthcare spending, and population growth and their relationship with life expectancy. Below you will find the two models that we will use as well as descriptions (Table 1), summary statistics (Table 2), and correlations (Table 3) of the explanatory variables that we will be examining.

Single Regression Model

$$\text{lifeexp} = \beta_0 + \beta_1 \log(\text{gdpcapita}) + u$$

Multiple Regression Model

$$\text{lifeexp} = \beta_0 + \beta_1 \log(\text{gdpcapita}) + \beta_2(\text{eduspending}) + \beta_3(\text{healthexp}) + \beta_4(\text{popgrowth}) + u$$

Table 1: Description of Variables

Variable	Description
Life Expectancy	Life expectancy at birth, total (years)
GDP per Capita	GDP per capita (current US$)
Education Spending	Government expenditure on education (% of GDP)
Healthcare Expenditures	Current healthcare expenditure (% of GDP)
Population Growth	Population growth (annual %)

Table 2: Summary Statistics

Variable	# of Obs.	Mean	Std. Dev	Min	Max
Life Exp.	146	72.29	7.60	53.71	84.21
log(GDP/Capita)	146	3.79	.63	2.43	5.07
Edu. Spending	146	4.51	1.28	1.45	7.96
Healthcare Spending	146	6.62	2.57	2.14	16.06
Pop. Growth	146	1.29	1.15	-1.80	4.92

Table 3: Correlation of Regressors

Variables	log(GDP/Capita)	Edu.Spending	Healthcare Spending	Pop. Growth
log(GDP/Capita)	1.00			
Edu. Spending	0.32	1.00		
Healthcare Spending	0.46	0.38	1.00	
Pop. Growth	-0.57	-0.22	-0.37	1.00

Gauss Markov Assumptions

In order for us to justify our use of a multiple regression model, we must ensure that our data meets the following assumptions:

1. **The regression model is linear** with respect to its coefficients and error term.

Our model satisfies the first Gauss Markov Assumption of linearity as all coefficients in the model are constants to be multiplied with an explanatory variable (i.e., it fits the linear form of $Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_n X_n + u$).

2. **Random Sampling**

We can assume that our data meets the condition of being randomly sampled from the population as it was collected by the World Bank, and thus satisfies the second assumption of random sampling.

3. **Non-Collinearity**

As shown in Table 3, none of our explanatory variables are perfectly correlated to another, thus our data meets the third assumption of non-collinearity.

4. **Zero Conditional Mean**

As shown in Figure 2 of the appendix, a plot of the residuals of our multiple linear regression model follows a normal distribution with a mean of zero, and thus our data satisfies the fourth assumption.

5. **Homoscedasticity**

The fifth assumption is that our error term has a constant variance (no heteroscedasticity). As can be seen in Figure 3 of the appendix (residuals vs. fitted plot) our residuals have a relatively constant variance with the exception of a few outliers, and thus we satisfy the fifth Gauss Markov Assumption.

6. **No Multicollinearity**

The final assumption is that none of our independent variables are perfect linear functions of any of our other explanatory variables (i.e., perfect correlation coefficient of +1 or -1). As can be seen in Table 3, none of our variables are perfectly correlated to any of our other variables, and thus we satisfy the sixth assumption.

Since our data meets the six assumptions stated above, we have justified that the best linear unbiased estimators of the coefficients in our model will be given by the OLS linear regression model.

Results

As discussed in the data section we used the following models to examine the relationship between our explanatory variables and life expectancy:

Single Regression Model

$$lifeexp = \beta_0 + \beta_1 log(gdpcapita) + u$$

Result: $lifeexp = 31.47 + 10.78log(gdpcapita)$

$N = 146$, Adj. $R^2 = .81$

Multiple Regression Model

$$lifeexp = \beta_0 + \beta_1 log(gdpcapita) + \beta_2(eduspending) + \beta_3(healthexp) + \beta_4(popgrowth) + u$$

Result: $lifeexp = 36.02 + 9.67log(gdpcapita) - 0.16(eduspending) + 0.22(healthexp) - 0.81(popgrowth)$

$N = 145$, Adj. $R^2 = .82$

Table 4: Estimation Results

Indep. Variables	Model (1)	Model (2)
log(GDP/Capita)	10.78*** (.44)	9.67*** (.56)
Edu. Spending		-0.16 (.23)
Healthcare Spending		.22* (.12)
Pop. Growth		-0.81*** (.29)
Intercept	31.47*** (1.69)	36.02*** (2.27)
No. of Obs.	146	145
Adj. R^2	.81	.82

*Statistically significant at 10%, **5%, ***1%.

Interpretation

The results of our single linear regression model indicate that GDP per capita is a decent predictor of life expectancy. In a sample of 146 countries and regions our linear regression model of life expectancy and log(gdpcapita) yielded an adjusted R^2 of .81, indicating that our model was fairly strong at explaining the relationship between national income and life expectancy. The coefficient of log(gdpcapita) was 10.78, signifying that for every 1 unit increase in log(gdpcapita), which corresponds to a tenfold increase in GDP per capita, a country can expect to see the life expectancy of its citizens increase by 10.78 years. Intuitively it makes sense that the relationship between GDP per capita and life expectancy is logarithmic in nature, as life expectancy will increase by a diminishing amount as national income increases. Adding education spending to create our first multiple linear regression model had no effect on the predictive accuracy of life expectancy, as the adjusted R^2 of this model was the same as our single linear regression model of life expectancy and GDP per capita. Additionally, our coefficient for education spending was not found to be statistically significant at even the 10% level and therefore we cannot say for certain that the coefficient is significantly different than zero.

Our second multiple regression model included healthcare spending and population growth as additional explanatory variables. The model yielded an adjusted R^2 of .82, only marginally higher than the adjusted R^2 that our single linear regression model yielded of .81. One thing to take away from the results of our second multiple linear regression model is the effect of population growth on life expectancy. Unsurprisingly, an increase in population is associated with a decrease in life expectancy. The coefficient of population growth is -0.81, indicating that every unit increase of 1% in population growth is associated with a decrease in life expectancy by 0.81 years. Additionally, the coefficient of population growth is the only other coefficient in our model besides the coefficient of log(gdpcapita) to be found statistically significant at the 1% level.

Statistical Inference

Table 5: Statistical Significance of Variables (Multiple Regression Model)

Indep. Variables	t-statistic	P-value (95% Confidence Int.)
log(GDP/Capita)	17.37	0.000
Education Spending	-.068	0.497
Healthcare Spending	1.78	0.077
Pop. Growth	-2.80	0.006

We can examine the statistical significance of each variable by looking at their respective t-statistic values and p-values (listed in Table 5). It's evident that national income is the most statistically significant of our explanatory variables given it's large t-statistic and significance at the 1% level, which is what we had hoped to see. The relatively large coefficient of log(gdpcapita) in both of our models also indicates the strong relationship between national income and a country's life expectancy. Population growth was the second most statistically significant explanatory variable, as it was also found to be significant at the 1% level. While this validates the hypothesis that population growth is correlated to life expectancy, the magnitude of its coefficient indicates that its effect is small in comparison to that of national income.

One of the more interesting takeaways from the data in Table 5 was the lack of statistical significance of the coefficient for education spending. As discussed briefly in the literature review section, a statistical relationship has been found between an individual's education level and their life expectancy. Our model, however, indicates that this relationship does not necessarily carry over to the national level (i.e., no significant relationship between national income and national life expectancy).

Robustness Test

Due to the relative insignificance of several of our explanatory variables we performed an F-test to test the hypothesis that the simpler of our two models is a better fit for our data set. Given that neither the coefficient of education spending or healthcare spending was found to be significant at the 5% level we will remove these variables to form our restricted model:

$$\text{lifeexp} = \text{lifeexp} = \beta_0 + \beta_1 \log(\text{gdpcapita}) + \beta_2(\text{popgrowth}) + u$$

Performing an F-test yielded us an F-value of 1.61, which was smaller than our critical value of 2.30. We can therefore accept the null hypothesis and conclude that education spending and healthcare spending are jointly insignificant in our model. More generally, we can now state that a multiple regression model of life expectancy, GDP per capita, and population growth is a better fit for our data than the original multiple regression model used.

Conclusion

Our initial hypothesis was that national income was positively correlated to life expectancy. Instinctively this made sense, as countries with higher levels of national income likely have better healthcare infrastructure and its citizens have higher standards of living, among other factors that have been shown to increase life expectancy. Based on the results of our regression models we can conclude that this hypothesis is correct. There is a statistically significant relationship between national income and life expectancy. More specifically, for every tenfold increase in GDP per capita a country can expect the life expectancy of its citizens to increase by approximately 11 years. Additionally, we found that there is a statistically significant relationship between population growth and life expectancy, although it's effect is much smaller than that of national income. For every unit increase of 1% in population growth one can expect to see the average life expectancy of a country to fall by 0.81 years. Instinctively this also makes sense, as citizens of highly populated countries may experience a lower standard of living as a result of crowding for example.

We were surprised by the statistical insignificance of the relationship between education spending and life expectancy. As briefly discussed at the beginning of the paper, a relationship has been found between an individual's education level and their life expectancy. This relationship, however, does not seem to transfer to the national level.

Although it was not examined in this paper, we further believe that the relationship between national income and life expectancy is causal. Countries that are able to produce more with a smaller population (the result being a higher GDP per capita) are able to improve their healthcare infrastructure and its citizens are likely to increase their standard of living overtime in comparison to citizens of poorer countries. One way to examine the causality of the relationship between national income and life expectancy would be to analyze changes in a country's national income and average life expectancy over time. Unfortunately we were not able to perform this analysis, as the data collected by the World Bank on GDP per capita and life expectancy is relatively incomplete until nearly 10 years ago. We hope that our single time period analysis of the relationship between income and life expectancy will motivate new research into the factors that can have an effect on life expectancy over time.

References

Chetty, Raj et al. “The Association Between Income and Life Expectancy in the United States, 2001-2014.” *JAMA* vol. 315,16 (2016): 1750-66. doi:10.1001/jama.2016.4226

“Current health expenditure (% of GDP).” *The World Bank,* data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS

“GDP per capita (Current US$).” *The World Bank*, data.worldbank.org/indicator/NY.GDP.PCAP.CD

“Government Expenditure on Education, Total (% of GDP).” *The World Bank*, data.worldbank.org/indicator/SE.XPD.TOTL.GD.ZS

Hummer, Robert A, and Elaine M Hernandez. “The Effect of Educational Attainment on Adult Mortality in the United States.” *Population bulletin* vol. 68,1 (2013): 1-16.

“Life Expectancy at Birth, Total (Years).” *The World Bank*, data.worldbank.org/indicator/SP.DYN.LE00.IN

“Population growth (annual %).” *The World Bank,* data.worldbank.org/indicator/SP.POP.GROW

Roser, Max, et al. “Life Expectancy.” *Our World in Data*, 23 May 2013, ourworldindata.org/life-expectancy

Twomey, Theresa Chalhoub and Madeline. “Income Inequality and Life Expectancy.” *Center for American Progress*, www.americanprogress.org/issues/healthcare/reports/2018/08/06/454149/income-inequality-life-expectancy/.

Appendix

Figure 1: Predicted vs. Actual Plot (Single Regression Model)

Figure 2: Histogram of Residuals (Single Regression Model)

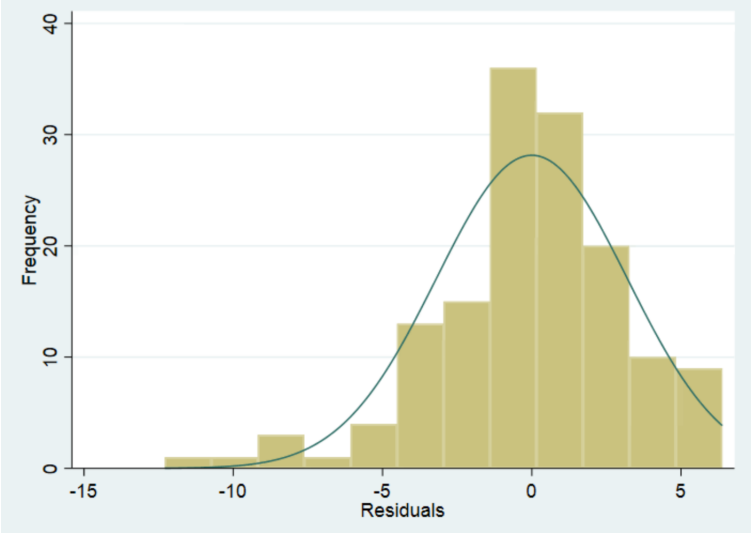

Figure 3: Residuals vs. Fitted Plot (Single Regression Model)